

ARTICLE

Received 25 Nov 2015 | Accepted 13 Jun 2016 | Published 13 Jul 2016

DOI: 10.1038/ncomms12209

OPEN

A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells

Ailone Tichon¹, Noa Gil¹, Yoav Lubelsky¹, Tal Havkin Solomon², Doron Lemze³, Shalev Itzkovitz³, Noam Stern-Ginossar² & Igor Ulitsky¹

Thousands of long noncoding RNA (lncRNA) genes are encoded in the human genome, and hundreds of them are evolutionarily conserved, but their functions and modes of action remain largely obscure. Particularly enigmatic lncRNAs are those that are exported to the cytoplasm, including NORAD—an abundant and highly conserved cytoplasmic lncRNA. Here we show that most of the sequence of NORAD is comprised of repetitive units that together contain at least 17 functional binding sites for the two mammalian Pumilio homologues. Through binding to PUM1 and PUM2, NORAD modulates the mRNA levels of their targets, which are enriched for genes involved in chromosome segregation during cell division. Our results suggest that some cytoplasmic lncRNAs function by modulating the activities of RNA-binding proteins, an activity which positions them at key junctions of cellular signalling pathways.

¹Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. ²Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. ³Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. Correspondence and requests for materials should be addressed to I.U. (email: igor.ulitsky@weizmann.ac.il).

Genomic studies conducted over the past 15 years have uncovered the intriguing complexity of the transcriptome and the existence of tens of thousands of long noncoding RNA (lncRNA) genes in the human genome, which are processed similarly to mRNAs but appear not to give rise to functional proteins¹. While some lncRNA genes overlap other genes and may be related to their biology, many do not, and these are referred to as long intervening noncoding RNAs, or lincRNAs. An increasing number of lincRNAs are implicated in a variety of cellular functions, and many are differentially expressed or otherwise altered in various instances of human disease²; therefore, there is an increasing need to decipher their modes of action. Mechanistically, most lincRNAs remain poorly characterized, and the few well-studied examples consist of lincRNAs that act in the nucleus to regulate the activity of loci found in *cis* to their sites of transcription³. These include the XIST lincRNA, a key component of the X-inactivation pathway, and lincRNAs that are instrumental for imprinting processes, such as AIRN⁴. However, a major portion of lincRNAs are exported to the cytoplasm: indeed, some estimates based on sequencing of RNA from various cellular compartments suggest that most well-expressed lincRNAs are in fact predominantly cytoplasmic¹.

The functional importance and modes of action of cytoplasmic lincRNAs remain particularly poorly understood. Some lincRNAs that are transcribed from regions overlapping the start codons of protein-coding genes in the antisense orientation can bind to and modulate the translation of those overlapping mRNAs⁵, and others have been proposed to pair with target genes through shared transposable elements found in opposing orientations⁶. Two lincRNAs that are spliced into circular forms were shown to act in the cytoplasm by binding Argonaute proteins (in one case, through ~70 binding sites for a miR-7 microRNA⁷) and act as sponges that modulate microRNA-mediated repression^{7,8}. Such examples are probably rare, as few circRNAs and few lincRNAs contain multiple canonical microRNA-binding sites⁹. It is not clear whether other cytoplasmic lincRNAs can act as decoys for additional RNA-binding proteins through a similar mechanism of offering abundant binding sites for the factors.

The Pumilio family consists of highly conserved proteins that serve as regulators of expression and translation of mRNAs that contain the Pumilio recognition element (PRE) in their 3'-untranslated regions (3'-UTRs)¹⁰. Pumilio proteins are members of the PUF family of proteins that is conserved from yeast to animals and plants, and whose members repress gene expression either by recruiting 3' deadenylation factors and antagonizing translation induction by the poly(A) binding protein¹¹, or by destabilizing the 5' cap-binding complex. The *Drosophila* Pumilio protein is essential for proper embryogenesis, establishment of the posterior-anterior gradient in the early embryo, and stem cell maintenance¹². Related roles were observed in other invertebrates¹⁰, and additional potential functions were reported in neuronal cells¹³. There are two Pumilio proteins in humans, PUM1 and PUM2 (ref. 10), which exhibit 91% similarity in their RNA-binding domains, and which were reported to regulate a highly overlapping but not identical set of targets in HeLa cells¹⁴. Mammalian Pumilio proteins have been suggested to be functionally important in neuronal activity¹⁵, ERK signalling¹⁶, germ cell development¹⁷ and stress response¹⁵. Therefore, modulation of Pumilio regulation is expected to have a significant impact on a variety of crucial biological processes.

Here, we characterize NORAD—an abundant lincRNA with highly expressed sequence homologues found throughout placental mammals. We show that NORAD is bound by both PUM1 and PUM2 through at least 17 functional binding sites. By perturbing NORAD levels in osteosarcoma U2OS cells, we show that NORAD modulates the mRNA abundance of

Pumilio targets, in particular those involved in mitotic progression. Further, using a luciferase reporter system we show that this modulation depends on the canonical Pumilio binding sites.

Results

NORAD is a cytoplasmic lincRNA conserved in mammals. In our studies of mammalian lincRNA conservation, we identified a conserved and abundant lincRNA currently annotated as *LINC00657* in human and *2900097C17Rik* in mouse, and recently denoted as 'noncoding RNA activated by DNA damage' or *NORAD*¹⁸. *NORAD* produces a 5.3 kb transcript that does not overlap other genes (Fig. 1a), starts from a single strong promoter overlapping a CpG island, terminates with a single major canonical poly(A) site, but is unspliced, unlike most long RNAs (Fig. 1b). Similar transcripts with substantial sequence homology can be seen in EST and RNA-seq data from mouse, rat, rabbit, dog, cow, and elephant. *NORAD* does not appear to be present in opossum, where a syntenic region can be unambiguously identified based on both flanking genes with no evidence of a transcribed gene in between them, and no homologues could be found in more basal vertebrates. *NORAD* is ubiquitously expressed across tissues and cell lines in human, mouse and dog, with comparable levels across most embryonic and adult tissues (Supplementary Fig. 1) with the exception of neuronal tissues, where *NORAD* is more highly expressed. In the presently most comprehensive data set of gene expression in normal human tissues, compiled by the GTEX project (<http://www.gtexportal.org/>), the 10 tissues with the highest *NORAD* expression all correspond to different regions of the brain (highest level in the frontal cortex with a reads per kilobase per million reads (RPKM) score of 142), with levels in other tissues varying between an RPKM of 78 (pituitary) to 27 (pancreas). Comparable levels were also observed across ENCODE cell lines, with the highest expression in the neuroblastoma SK-N-SH cells (Fig. 1d). The high expression levels of *NORAD* in the germ cells have probably contributed to the large number of closely related *NORAD* pseudogenes found throughout mammalian genomes. There are four pseudogenes in human that share >90% homology with *NORAD* over >4 kb, but they do not appear to be expressed, with the notable exception of HCG11, which is annotated as a lincRNA and is expressed in a variety of tissues but at levels ~20-times lower than *NORAD* (based on GTEX and ENCODE data, Fig. 1d). Because of this difference in expression levels, we assume that while most of the experimental methods we used are not able to distinguish between *NORAD* and HCG11, the described effects likely stem from the *NORAD* locus and not from HCG11. Using single-molecule *in situ* hybridization (smFISH)¹⁹ in U2OS cells, we found that *NORAD* localizes almost exclusively to the cytoplasm (Fig. 1c and Supplementary Fig. 2) and similar cytoplasmic enrichment is observed in other cells lines (Fig. 1d). The number of *NORAD* copies expressed in a cell is ~80 based on the RPKM data (assuming an RPKM of 1 roughly corresponds to a single copy per cell) and 68 ± 8 based on the smFISH experiments that we have performed on U2OS cells, with 94% of *NORAD* copies located in the cytoplasm and 6% in the nucleus.

NORAD is a bona fide noncoding RNA. *NORAD* is computationally predicted to be a noncoding RNA by the PhyloCSF (Fig. 1e) and Pfam/HMMER pipelines²⁰, with CPAT²¹ and CPC²² giving it borderline scores due to the presence of an open reading frame (ORF) with >100aa (see below) and similarity to hypothetical proteins (encoded by *NORAD* homologues) in other primates. Therefore, we also examined whether *NORAD* contains any translated ORFs using Ribo-seq data²³. When examining

ribosome footprinting data sets from diverse human cell lines (MDA-MB-231 (ref. 24), HEK-293 (ref. 25), U2OS²⁶, and KOPT-K1 (ref. 27)), we did not observe any substantial footprints over any of the ORFs in NORAD, including a poorly conserved 108 aa ORF found close to the 5'-end of the human transcript (Fig. 1e). Interestingly, substantial pileups of ribosome-protected fragments were observed at the very 5'-end of NORAD in all Ribo-seq data sets we examined (Fig. 1e and Supplementary Fig. 3), but those did not overlap any ORFs with either the canonical AUG start codon or any of the common alternative start codons (Supplementary Fig. 3), nor did they encode any conserved amino acid stretches in any of the frames. We conclude that it is highly unlikely that NORAD is translated into a functional protein under regular growth conditions in those cell types, and the footprints observed in Ribo-seq data result from either a ribosome stalled at the very beginning of a transcript, or from a contaminant footprint of a different ribonucleoprotein complex, as such footprints are occasionally present in Ribo-seq experiments^{25,28}. It remains possible that NORAD is translated in other conditions and contexts.

NORAD contains at least 12 structured repeated units. When comparing the NORAD sequence to itself, we noticed a remarkable similarity among some parts of its sequence (Fig. 2a). Manual comparison of the sequences revealed that the central ~3.5 kb of NORAD in human, mouse, and other mammalian species can be decomposed into 12 repeating units of ~300 nt each. Interestingly, these units appear to have resulted from a tandem sequence duplication that occurred at least 100 million years ago, before the split of the eutherian mammals, as when performing pairwise comparisons, units from different species were more similar to each other than to other units from the same species. Overall, the sequences have diverged to a level where there are no sequence stretches that are strictly identical among all the repeats in human. At the core of the most conserved regions within the repeats we identify four sequence and structure motifs (Fig. 2d,e), some combination of which appears in each of the repeats 1–10: (i) one or two PREs, defined by the consensus UGURUAUA; (ii) a short predicted stem-loop structure with four paired bases and a variable loop sequence. The importance of the structure is supported by the preferential A→G and

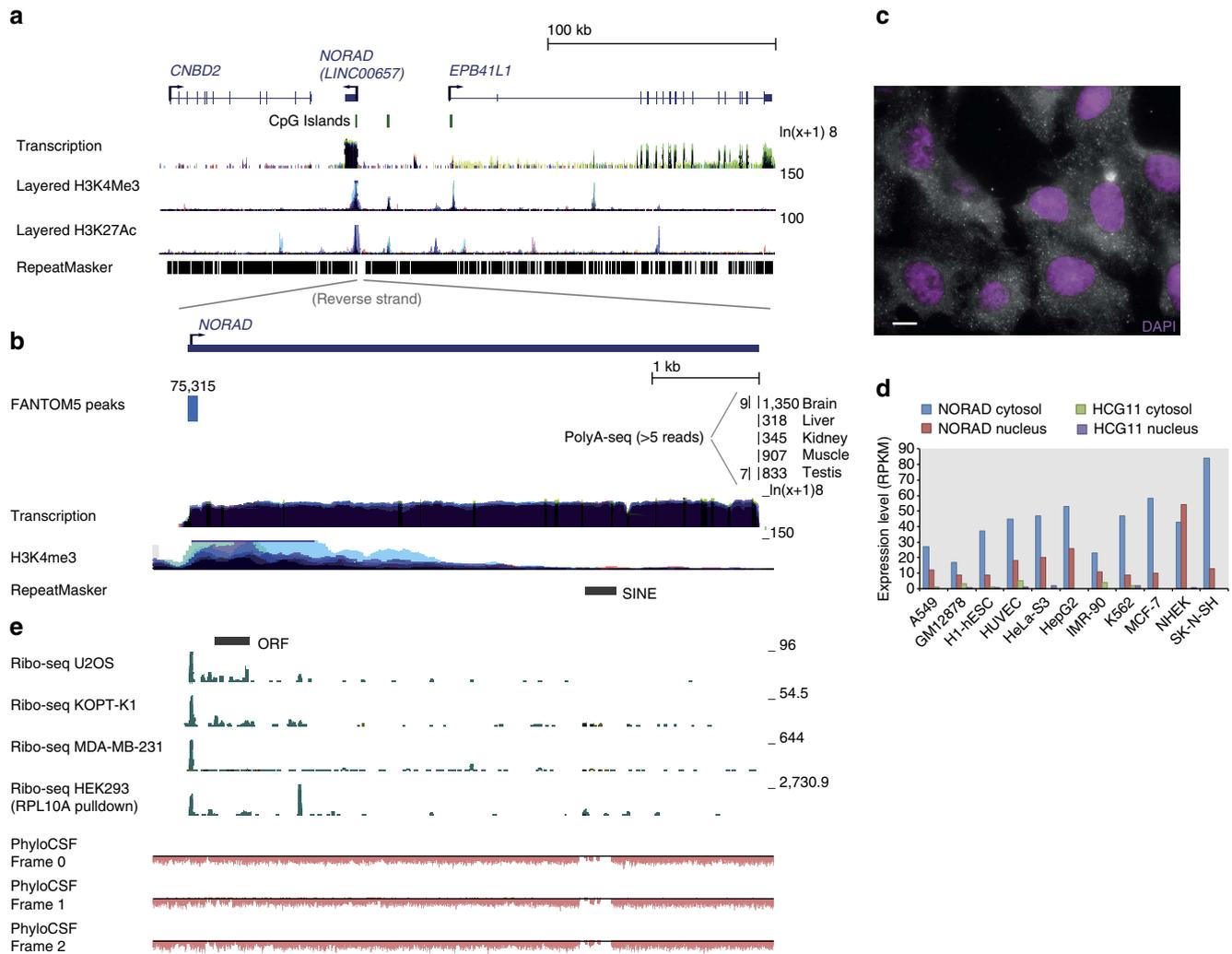


Figure 1 | Overview of the human NORAD locus. (a) Genomic neighbourhood of NORAD. CpG island annotations and genomic data from the ENCODE project taken from the UCSC genome browser. **(b)** Support for the NORAD transcription unit. Transcription start site information taken from the FANTOM5 project⁴⁵. Polyadenylation sites taken from PolyA-seq data set⁴⁶. ENCODE data sets and repeat annotations from the UCSC browser. **(c)** Predominantly cytoplasmic localization of NORAD by smFISH. Scale bar, 10 μm. See Supplementary Fig. 2 for RNA-FISH following NORAD knockdown. **(d)** Expression levels of NORAD and HCG11 in the ENCODE cell lines (taken from the EMBL-EBI Expression Atlas (<https://www.ebi.ac.uk/gxa/home>)). **(e)** Support for the noncoding nature of NORAD. Ribosome-protected fragments from various human cell lines (MDA-MB-231 (ref. 24), HEK-293 (ref. 25), U2OS²⁶ and KOPT-K1 (ref. 27)) mapped to the NORAD locus as well as PhyloCSF⁴⁷ scores. All PhyloCSF scores in the locus are negative.

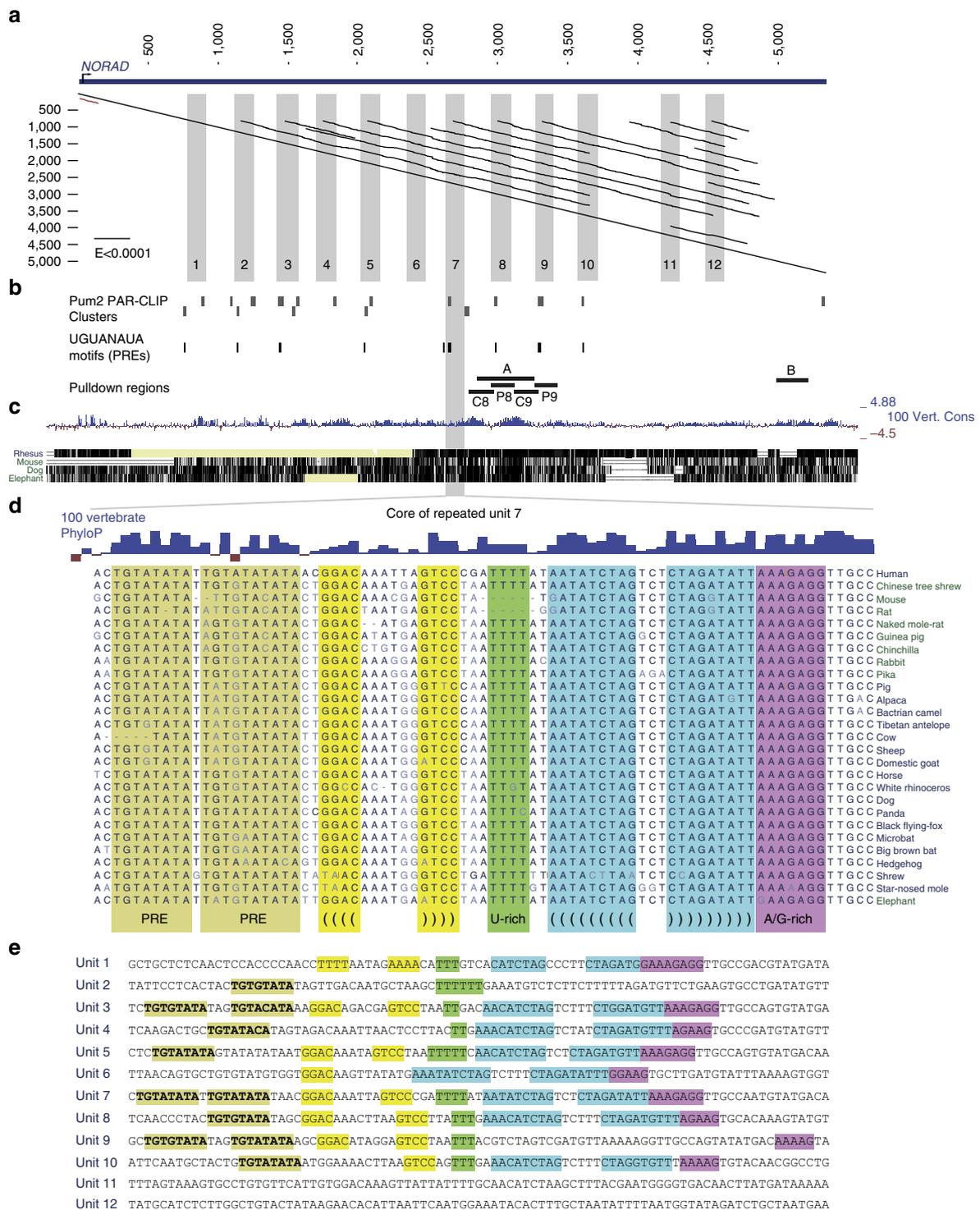


Figure 2 | The repeated nature of the NORAD RNA. (a) A dotplot computed using plalign⁴⁸ (<http://fasta.bioch.virginia.edu/>) comparing NORAD with itself. The off-diagonal lines indicate high scoring local alignments between different parts of the sequence. Grey boxes indicate the core of the 12 manually annotated repeated units. **(b)** Clusters identified by PARalyzer⁴⁹ within the NORAD sequence using the PUM2 PAR-CLIP data³⁰, positions of PRE UGURUAUA motifs, and regions used for *in vitro* transcription and pull-down of NORAD fragments. **(c)** Sequence conservation of the NORAD locus, with PhyloP⁵⁰ scores for single-base level conservation. **(d)** Detailed conservation of the seventh repeated unit. Shaded regions indicate the five motifs present in most repeated units. **(e)** Core sequences of the human repeated units, with the same shading as in **d**.

G → A mutations in the second stem-loop that would preserve the stem (Fig. 2d and Supplementary Fig. 4, also detected by EvoFold²⁹); (iii) a U-rich stretch of 2–5 bases; and (iv) a stem-loop structure with eight or nine predicted base pairs.

Further sequence conservation is found upstream and downstream of these motifs. Interestingly, the sequences of some of the repeated units, namely 3–5 and 7–9, appear to be more constrained during mammalian evolution than others

(Fig. 2c), and those units also tend to contain most of the repeat motifs, with more intact sequences and structures (Fig. 2e).

NORAD contains multiple functional Pumilio binding sites.

To identify potential protein binding partners of the repeating units and of other NORAD fragments, we first amplified the eighth repeat unit and a region from the 3'-end of NORAD (regions A and B marked in Fig. 2b), transcribed them *in vitro* in the sense (regions A and B) and antisense (region A) orientations using the T7 polymerase with biotinylated UTP bases, incubated the labelled RNA with U2OS cell lysate, and subjected the resulting pulldown material to mass spectrometry. Among the proteins identified as binding different regions of NORAD (Supplementary Data 1) we focus here on two that have predicted binding sites within the repeat units—PUM1 and PUM2, the two vertebrate Pumilio proteins¹⁰. PUM1 and PUM2 proteins were enriched when we performed similar pulldowns followed by western blots using the PRE-containing regions within repeats 8 and 9 (P8 and P9 in Fig. 2b, Fig. 3a and Supplementary Fig. 5) relative to the adjacent sequences (C8 and C9 marked in Figs 2b and 3a and Supplementary Fig. 5). In addition, enrichment was strongly reduced when one of the two PREs in region P9 had been mutated (Fig. 3b and Supplementary Fig. 5). To gain additional support for a direct interaction between PUM2 and NORAD, we reanalyzed PAR-CLIP data from HEK-293 cells³⁰ and found that PUM2 binds at least 17 sites on NORAD (Fig. 2b). These experimentally verified sites (all exhibiting T→C mutations characteristic of PAR-CLIP) overlapped 10 out of the 11 PREs within repeated units 2–10. It is notable that NORAD has an unusual density of PREs encoded in its sequence—there are 17 non-overlapping instances of the UGURUAUA motifs in NORAD compared with 0.38 expected by chance ($P < 0.001$, see Methods). The number and density of Pumilio motifs within NORAD are higher than those found in all but one human gene (*PLCXD1*, which has 18 PREs mostly located in transposable elements, compared with 0.12 expected).

To test whether NORAD also co-precipitates with PUM1 and PUM2 in U2OS cells, we performed RNA immunoprecipitation (RIP) of both proteins followed by quantitative real-time PCR, and found a striking enrichment of the NORAD transcript, with a stronger enrichment observed for PUM2 (Methods and Fig. 3c,d). We conclude that NORAD contains at least 17 confident binding sites for Pumilio proteins, most of which appear in conserved positions within the repeated units. Surprisingly, despite the presence of a large number of binding sites, NORAD was not more susceptible to change following PUM1 and PUM2 overexpression or knockdown in U2OS cells (see below) than targets with few PREs, suggesting that NORAD is resistant to substantial degradation by the Pumilio proteins under the tested conditions.

With ~70 NORAD transcripts per cell (Fig. 1c) and at least 17 functional PREs (Fig. 2b), NORAD possesses the capacity to simultaneously bind ~1,200 PUM proteins. Quantitative western blot analysis comparing U2OS cell lysates to recombinant proteins expressed in bacteria revealed that PUM1 and PUM2 are expressed at ~200 and ~550 copies per cell, respectively (Supplementary Fig. 6). The sites offered by NORAD for Pumilio protein binding, as well as the potential interactions made possible between Pumilio proteins and other NORAD-interacting factors when bound simultaneously, can be sufficient for eliciting a significant effect on the number of functional Pumilio proteins that are available to act as repressors of their other targets.

NORAD perturbations preferentially affect Pumilio targets.

As PUM1 and PUM2 are reported to affect mRNA stability^{11,31}, we next tested whether changes in NORAD expression affect the levels of Pumilio targets. We defined Pumilio target genes as those having at least two extra UGUANAUA sites in their 3'-UTRs over the number of sites expected given the 3'-UTR length of the transcripts. To validate that such genes indeed represent Pumilio targets, we knocked down (KD, Supplementary Fig. 7) and overexpressed (OE) PUM1 and PUM2 separately in U2OS cells

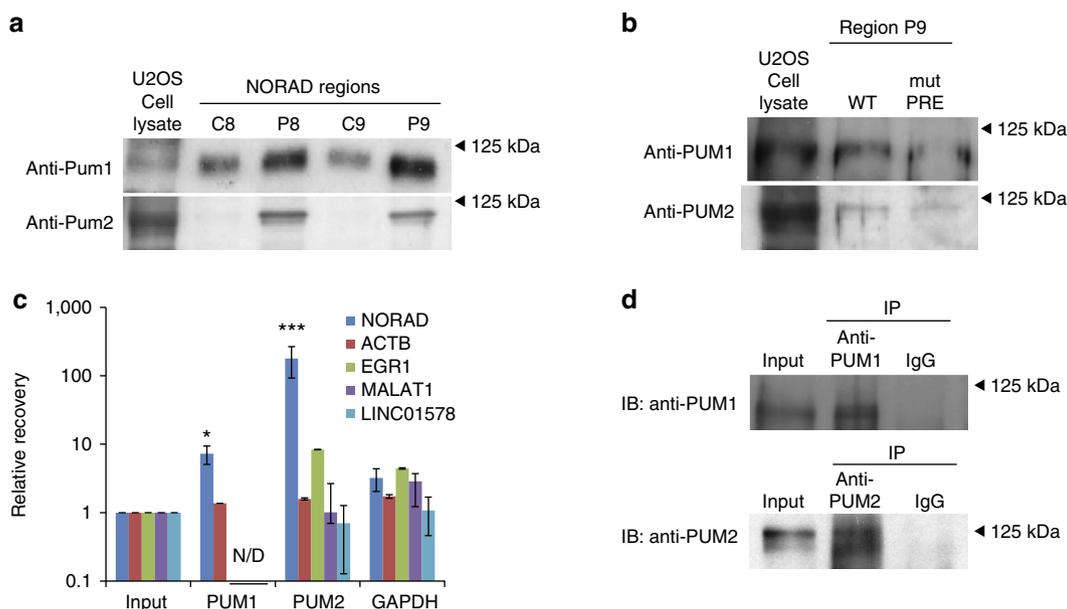


Figure 3 | Pumilio proteins bind NORAD. (a) Western blots for PUM1 and PUM2 following pulldowns using the indicated *in vitro* transcribed regions (marked in Fig. 2b). (b) Western blots for PUM1 and PUM2 following pulldowns using *in vitro* transcribed RNAs from synthetic oligos with WT or mutated PRE. (c) Recovery of the indicated transcripts in the input and in the indicated IPs. All enrichments are normalized to GAPDH mRNA and to the input sample as described in Methods. * $P < 0.05$, *** $P < 0.001$ (Tukey's HSD test). Error bars represent s.e.m. based on at least three independent replicates. (d) Western blots of the indicated factors in the input and IP samples. HSD, honest significant difference; WT, wild type.

and observed significant upregulation and downregulation of predicted *Pumilio* targets following KD and OE, respectively (Fig. 4a,b). NORAD was then perturbed using either one of two individual siRNAs (siRNA 1 and siRNA 2, Supplementary Fig. 8A) or a pool of four siRNAs (Dharmacon), with the pool yielding ~4-fold knockdown and individual siRNAs yielding ~2-fold knockdown (Supplementary Fig. 8A). We obtained consistent effects with the two siRNAs 48 h after transfection (Supplementary Fig. 8B, Supplementary Data 2), with 51 genes consistently downregulated by at least 20% and 23 genes consistently upregulated by at least 20% after treatment with both siRNAs. The stronger knockdown using a pool of siRNAs (Supplementary Fig. 8A) resulted in more substantial changes in gene expression—584 genes were consistently downregulated by

at least 30% in two replicates, and 68 genes were consistently upregulated (Supplementary Data 2). To test the consequences of increased NORAD levels, we cloned NORAD into an expression vector, where its transcription was driven by a CMV promoter, and transfected this vector into U2OS and HeLa cells, which resulted in 2–16-fold NORAD upregulation. Changes following NORAD downregulation at 24 h were strongly inversely correlated with the changes observed 24 h after NORAD OE (Supplementary Fig. 8C and Supplementary Data 2, Spearman $r = -0.54$, $P < 10^{-10}$), suggesting that the differential expression was indeed driven by changes in NORAD abundance. Strikingly, *Pumilio* targets were repressed more than controls when NORAD was downregulated, and their expression levels increased more than controls when NORAD was upregulated in both U2OS and

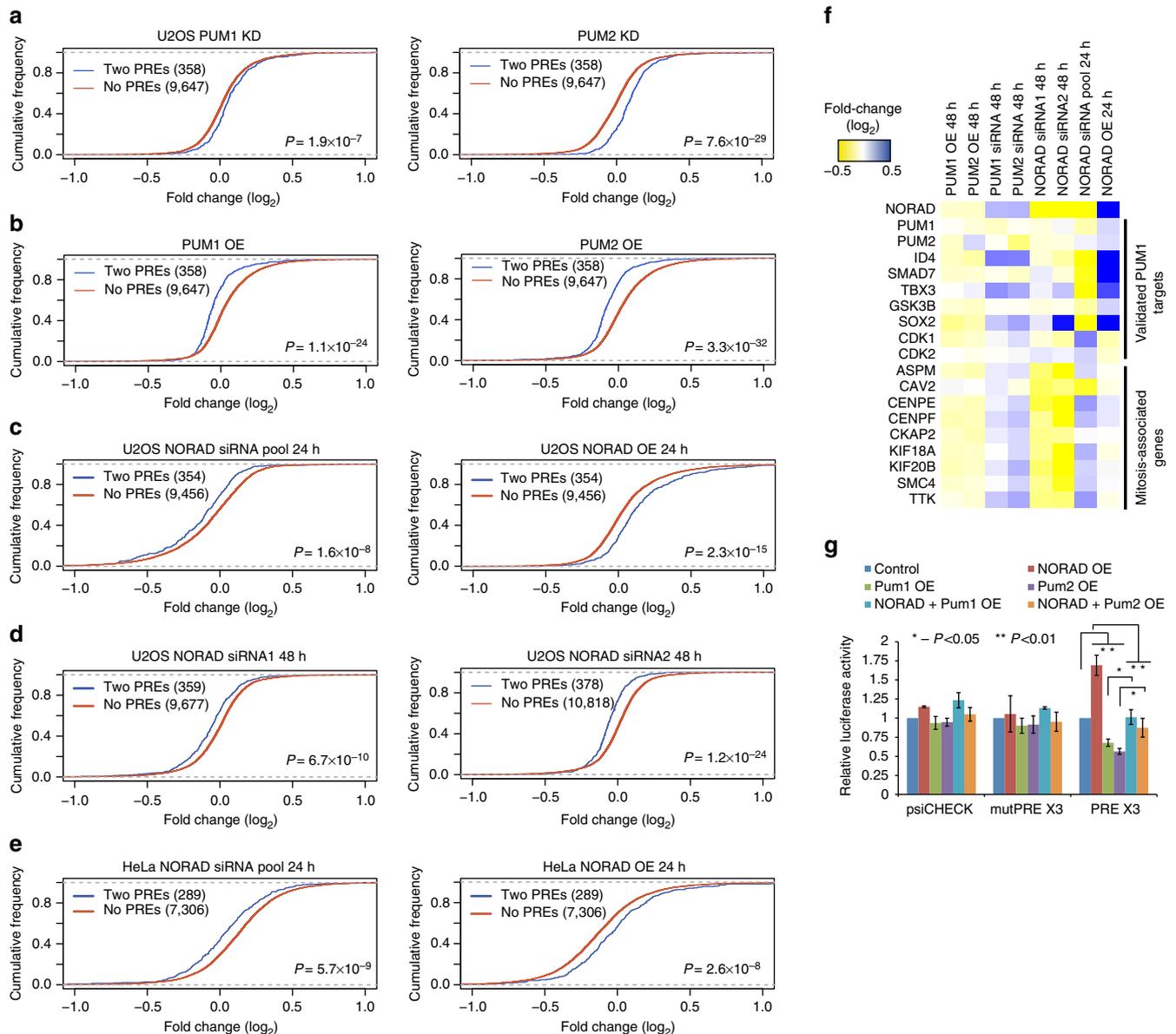


Figure 4 | NORAD modulates expression of *Pumilio* targets. (a–e) Changes in expression of *Pumilio* targets compared with controls, following the indicated treatment. Numbers indicate the number of genes in each group that were sufficiently expressed (see Methods). ‘2 PREs’ are genes that contain at least two canonical PREs over what is expected by chance in their 3′-UTRs, ‘controls’ are those genes that do not contain more sites in their 3′-UTRs than expected by chance. (f) Changes in expression of NORAD, PUM1/2, validated targets of PUM1 (ref. 33) and genes with annotated roles in the M phase of the cell cycle and/or the mitotic spindle following the indicated perturbations. (g) Changes in the luciferase activity measured from the indicated vectors (top, psiCHECK is a control vector) and RNA expression measured using RT-PCR (bottom) following overexpression of the indicated genes and combinations. Error bars represent s.e.m. based on at least three independent replicates.

HeLa cells (Fig. 4c–e). These differences remained significant after controlling for the increased lengths of the 3′-UTRs of genes bearing Pumilio motifs (Supplementary Fig. 9A) and when considering genes with PUM2 PAR-CLIP clusters in their 3′-UTR as determined in HEK-293 cells (these effects were strongest 48 h after transfection, Supplementary Fig. 9B). Genes with multiple PREs were generally more affected than those with fewer sites (Supplementary Fig. 10A). Differences between Pumilio targets and controls were observed when considering exon-mapping and not when considering intron-mapping reads, pointing at post-transcriptional regulation³² (Supplementary Fig. 10B). Lastly, we observed consistent effects in validated PUM1 targets³³ expressed in U2OS cells (Fig. 4f). These results suggest that hundreds of genes regulated by the two Pumilio proteins are sensitive to NORAD levels, with increased NORAD amounts alleviating repression of Pumilio targets and decreased NORAD amounts increasing repression.

When we inspected the Gene Ontology annotations enriched in the different sets of genes responsive to NORAD perturbations, after correction for multiple testing using TANGO³⁴, the only significantly enriched group were genes bound by PUM2 in the PAR-CLIP data and downregulated 48 h after NORAD knockdown. These genes were enriched with categories associated with cell cycle and mitosis, including ‘M phase of the cell cycle’ (eight genes; $P=6.4 \times 10^{-6}$) and ‘Spindle’ (eight genes; $P=1.2 \times 10^{-7}$). Interestingly, these genes were not substantially affected at 24 h after NORAD knockdown or overexpression (Fig. 4f), and enrichments of NORAD targets were also significant when compared with all PUM2-bound targets, suggesting a cumulative, and perhaps cell cycle-dependent, effect of NORAD perturbation on Pumilio targeting of genes involved in mitosis. These results are consistent with the chromosomal instability and mitotic defects observed in other cell types following TALEN-mediated deletion of NORAD¹⁸.

As Pumilio proteins may affect translation in addition to their effects on mRNA stability, we evaluated the translational consequences of NORAD perturbation after 48 h using Ribo-seq³⁵. Consistent with the RNA-seq data, the number of translating ribosomes on Pumilio targets was reduced following NORAD KD (Supplementary Fig. 11A). However, when normalizing for changes in mRNA levels, translation efficiency of Pumilio targets did not appear to be preferentially affected (Supplementary Fig. 11B), suggesting that the main effects of NORAD on Pumilio targets are through effects on mRNA stability rather than translation. This observation is consistent with reports that the mechanism of action of Pumilio proteins is through interaction with deadenylation complexes^{11,31} that can first affect protein translation, but eventually results in mRNA decay.

NORAD regulation is dependent on the canonical PREs.

To test whether regulation of Pumilio targets depends on the presence of canonical PREs, we utilized a luciferase reporter vector containing three strong PREs as well as a control reporter with mutated sites, in which the three 5′-UGUACAUA-3′ motifs were mutated to 5′-ACAACATA-3′ (mutPRE)^{11,31}. As expected, overexpression of PUM1 or PUM2 proteins in U2OS cells led to reduced luciferase activity in a PRE-dependent manner (Fig. 4g). Overexpression of NORAD, on the other hand, alleviated the repression of the PRE-containing luciferase mRNA, without affecting mutPRE-containing mRNA. Simultaneous OE of NORAD and the Pumilio proteins abrogated both effects, an observation consistent with our model that the effect of NORAD on Pumilio targets is mediated through Pumilio proteins (Fig. 4g and Supplementary Fig. 12A). Knockdowns of NORAD or PUM1

or PUM2 failed to yield a consistent effect on luciferase activity (Supplementary Fig. 12B), possibly because of the limited knockdown efficiency using siRNAs (Supplementary Fig. 12C) or through feedback regulation of PUM1 or PUM2 on their own mRNA. Overall, these results indicate that the NORAD-dependent changes in abundance of Pumilio targets are likely mediated through canonical PREs.

Discussion

To our knowledge, NORAD comprises the first example of a lncRNA that contains multiple highly conserved consensus binding sites for an RNA-binding protein (RBP), and that is required for proper regulation of the RBP targets at physiological levels. One particularly interesting question that remains open is the functional importance and roles of the other conserved elements found in the NORAD repeats, and in particular the two predicted hairpin structures, as such conserved secondary structures are rarely detectable in lncRNAs¹. It is possible that these structural elements serve as binding sites for other RBPs, whose binding may either facilitate the binding of PUM1 and PUM2 to NORAD or affect PUM1 or PUM2 protein stability or activity. We note that while the overall number of binding sites offered by NORAD for PUM1 and PUM2 (~1,200) is comparable to the Pumilio abundance in U2OS cells, which we estimate at ~200 and ~550 copies per cell for each of PUM1 and PUM2, respectively (Supplementary Fig. 6), these sites are outnumbered by the sites present in other expressed mRNAs, and therefore it is possible that NORAD does not merely titrate Pumilio proteins away from their other targets but rather induces a change in their activity, potentially by serving as a scaffold for interaction of Pumilio proteins with other factors. Potentially interesting candidates for interacting with NORAD repeats that were identified in the mass spectrometry analysis are known RBPs, such as IGF2BP1/2/3, XRN2 and PABPN1. In addition, we observed that the interferon response pathway proteins IFIT1/2/3/5 and their downstream companion PKR could bind NORAD sequence. IFIT proteins were observed to bind the antisense of the NORAD eighth repeat unit, suggesting that they may recognize a structural element rather than a primary sequence within the repeat, whereas PUM1 and PUM2 bound only the sense sequence, consistent with their known sequence specificity. We were so far unable to substantiate interactions with IFIT1 and PKR by reciprocal pulldown experiments.

While this manuscript was under review, Mendell and colleagues described a role for NORAD and PUM2 in ensuring chromosomal segregation fidelity in various human cells¹⁸. Further studies will be required in order to uncover the full spectrum of physiological consequences of the regulation of Pumilio targets by NORAD, but the enrichment of cytokinesis-related genes among the Pumilio targets that are sensitive to NORAD levels suggests that NORAD may modulate regulation of chromosomal segregation during mitosis by Pumilio, and might even affect the conserved roles of Pumilio in regulating asymmetric cell divisions during embryonic development. An intriguing question is whether the relatively high levels of NORAD in U2OS cells correspond to a basal state, in which NORAD exerts a minimal effect on PUM1 and PUM2 that is increased when stimuli increase NORAD expression, or to a state where NORAD actively buffers substantial regulation by PUM1 and PUM2. Most results point to the former scenario, as relatively modest overexpression of NORAD resulted in stronger effects on Pumilio activity than its knockdown. Another possibility suggested by the enrichment of cell cycle regulated genes among the most prominent NORAD and Pumilio targets is that this regulation is cell cycle-dependent.

Methods

Cell culture. Human cell lines U2OS (osteosarcoma, obtained from the ATCC) and HeLa (cervical carcinoma, obtained from the ATCC) were routinely cultured in DMEM containing 10% fetal bovine serum and 100 U penicillin/0.1 mg ml⁻¹ streptomycin at 37 °C in a humidified incubator with 5% CO₂.

Plasmids and siRNAs. Plasmid transfections were performed using polyethyleneimine (PEI)³⁶ (PEI linear, Mr 25,000 from Polyscience Inc). To overexpress NORAD, the full transcript of the lincRNA was amplified from human genomic DNA (ATCC NCI-BL2126) using the primers 5'-TGCCAGCGCAGAGAACTGCC-3' (Fw) and 5'-GGCACTCGGGAGTGTCAGGTTTC-3' (Rev), and cloned into a ZeroBlunt TOPO vector (Invitrogen), and then subcloned into the pcDNA3.1(+) vector (Invitrogen). PUM1 and PUM2 were overexpressed using pEF-BOS vectors^{37,38} (a kind gift of Prof. Takashi Fujita). As controls in overexpression experiments, we used pBlueScript II KS+ (Stratagene). Plasmids were used in the amount of 0.1 µg per 100,000 cells in 24-well plates for 24 h before cells were harvested. The luciferase experiments employed the following plasmids: pGL4.13; psiCheck-1 containing 3X wild-type PRE, which is underlined in the following sequence, 5'-TTGTTGTGCGAAAATTGTACATAAGCCAA-3'; psiCheck-1 containing 3X mutated PREs: 5'-TTGTTGTGCGAAAATACAACATAAGCCAA-3' and psiCheck-1 with no PRE, all previously described^{11,31} (a kind gift of Dr Aaron Goldstrohm). pGL4.13 was used in the amount of 5 ng per 20,000 cells in 96-well plates, while the different psiCheck-1 plasmids were used in the amount of 15 ng per 20,000 cells in 96-well plates.

Gene knockdown was achieved using siRNAs directed against NORAD, PUM1 and PUM2 genes (all from Dharmacon, Supplementary Table 1), while as control we used the mammalian non-targeting siRNA (Lincode Non-targeting Pool, Dharmacon), at final concentration of 50 nM for 24 or 48 h before further experimental procedures. The transfections into U2OS cells were conducted using PEI.

siRNA transfection into HeLa cells were conducted using 100 nM siRNA and Dharmafect (Dharmacon) transfection reagent and using siRNA buffer only as a control, and transfection of pcDNA3.1-NORAD was into HeLa cells was performed using Lipofectamine 2,000.

Real-time PCR analysis of gene expression. Total RNA was isolated using TRI reagent (MRC), followed by reverse transcription using an equal mix of oligo dT and random primers (Quanta), according to the manufacturer's instructions. For determination of all genes levels real-time PCR was conducted using Fast SYBR qPCR mix (Life Technologies). The primer sets used for the different genes are listed in Supplementary Table 2. The assays contained 10–50 ng sample cDNA in a final volume of 10 µl and were run on AB quantitative real-time PCR system ViiA 7 (Applied Biosystems). All genes expression levels in the different treatments are represented relative to their relevant control (Δ Ct) and normalized to GAPDH gene levels ($\Delta\Delta$ Ct).

Fluorescent *in situ* hybridization. Probe libraries were designed according to Stellaris guidelines and synthesized by Stellaris as described in Raj *et al.*¹⁹. Libraries consisted of 48 probes 20 nt each, complementary to the NORAD sequence according to the Stellaris guidelines (Supplementary Table 3). Hybridizations were done overnight at 30 °C with Cy5 labelled probes at a final concentration of 0.1 ng µl⁻¹. DAPI dye (Inno-TRAIN Diagnostik GmbH) for nuclear staining was added during the washes. Images were taken with a Nikon Ti-E inverted fluorescence microscope equipped with a ×100 oil-immersion objective and a Photometrics Pixis 1,024 CCD camera using MetaMorph software (Molecular Devices, Downingtown, PA). The image-plane pixel dimension was 0.13 µm. Quantification was done on stacks of 4–12 optical sections with Z-spacing of 0.3 µm. Dots were automatically detected using a custom Matlab program, implementing algorithms described in Raj *et al.*¹⁹. Briefly, the dot stack images were first filtered with a three-dimensional Laplacian of Gaussian filter of size 15 pixels and standard deviation of 1.5 pixels. The number of connected components in binary thresholded images was then recorded for a uniform range of intensity thresholds and the threshold for which the number of components was least sensitive to threshold selection was used for dot detection. Automatic threshold selection was manually verified and corrected for errors. Background dots were detected according to size and by automatically identifying dots that appear in more than one channel (typically <1% of dots) and were removed.

RNA pulldown. Templates for *in vitro* transcription were generated by amplifying the desired sequences from cDNA or from synthetic oligos, adding the T7 promoter to the 5'-end for sense and 3'-end for the antisense sequence (see Supplementary Table 2 for primer sequences). In addition, protein pulldown was performed using an oligo with the sequence of repeat #9 (5'-GTCTGATTTCATTTCA TTTACTGTGCTGTGTATATAGTGTATATAAGCGGCATATAGGAGTCCCTAAT TTACGTTACTGCTGATGTTAAAAAGGTTGCCAGTATATGACAAAAGTAGAA TTAGTAAACTACTACATTTGAGTACACTTTGTGTTAAAAATTCATAGGGA)-3' and an oligo that contains a mutation in its PRE (5'-GTCTGATTTCATTTCA TGTGCTACATATATAGTGTATATAAGCGGCATATAGGAGTCCCTAATTTAC GTCTAGTTCGATGTTAAAAAGGTTGCCAGTATATGACAAAAGTAGAATTA

GTAAACTACTACATTGAGTACACTTTGTGTTAAAAATTCATAGGGA-3') using the primers from Supplementary Table 2. Biotinylated transcripts were produced using the MEGAscript T7 *in vitro* transcription reaction kit (Ambion) and Biotin RNA labelling mix (Roche). Template DNA was removed by treatment with DnaseI (Quanta). Cells were lysed in buffer containing 20 mM Tris•HCl pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, 2 mM DTT, 0.5% Na-Deoxycholate, 0.5% NP-40) for 15 min on ice. The extract was cleared by centrifugation at 21,130g at 4 °C for 20 min. Extract containing 0.5–2 mg of protein was incubated with 2–20 pmole of biotinylated transcripts. The pulldown products were analysed by mass spectrometry and western blots. For the mass spectrometry the formed RNA-protein complexes were precipitated by Streptavidin-sepharose high-performance beads (GE Healthcare). Recovered proteins were then resolved on a 4–12% Express Page gradient gel (GeneScript), visualized by silver staining. The entire lane was extracted and analysed using mass spectrometry analysis as described³⁷. Briefly, peptide fragments were separated using a Nanosep3KD micro centrifuge tube (Pall, USA) and MS measurements were performed using a nano-electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) instrument (Applied Biosystems, Foster City, CA). The spectra were searched against human database with the use of the MASCOT (Matrix Science, London, UK).

Alternatively, the recovered proteins were separated on a 10% SDS-polyacrylamide gel electrophoresis gel, and used for western blotting with anti-PUM1 or anti-PUM2 antibodies (Bethyl Laboratories; anti-Pum1: A300-201A; anti-Pum2 A300-202A). In addition, RNA was isolated using TRI reagent from equal portion of the different protein-RNA pulldown complexes. This RNA was analysed using RT-PCR for loading control.

RNA immunoprecipitation (RIP). Immunoprecipitation (IP) of endogenous ribonucleoprotein complexes from whole-cell extracts was performed as described by Yoon *et al.*³⁹. In brief, cells were incubated with lysis buffer (20 mM Tris•HCl at pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, 2 mM DTT, 0.5% Na-deoxycholate, 0.5% NP-40, complete protease inhibitor cocktail (Sigma) and 100unit per ml RNase inhibitor (EURx)) for 15 min on ice and centrifuged at 15,870g for 15 min at 4 °C. Part of the supernatants was saved as total cell lysate input. The rest, containing 1–2 mg protein extract, was incubated for 2–3 h at 4 °C in gentle rotation with protein A/G magnetic beads (GeneScript). The beads were pre-washed and coated with antibodies against GAPDH (SantaCruz SC-32233 Biotechnology, diluted 1:1,000), PUM1 and PUM2 (Bethyl,Laboratoris, A300-201A and A3000202A respectively, diluted 1:1,000) at 4 °C in gentle rotation overnight. As a negative control, we incubated the magnetic beads-antibodies complexes with lysis buffer. The beads were washed five times with lysis buffer, each time separated by magnetic force. The remaining mixture of magnetic beads-antibodies-protein-RNA complexes were separated as half were mixed with sample buffer and boiled at 95C for 5 min for further analysis by Western blot and the other half was incubated with 1 mg ml⁻¹ Proteinase K for 30 min at 37C with gentle shaking to remove proteins. The remaining RNA was extracted by TRI reagent. The RNAs isolated from the IP materials were further assessed by RT-qPCR analysis as follows: IP material $\frac{\text{Examined gene levels}}{\text{GAPDH levels}} / \text{Total cell lysate} \frac{\text{Examined gene levels}}{\text{GAPDH levels}}$. Western blot was used in order to verify that the desired protein was indeed precipitated.

Ribosome profiling. U2OS cells, transfected with siRNAs, were lysed in lysis buffer (20 mM Tris•HCl pH 7.5, 150 mM KCl, 5 mM MgCl₂, 1 mM dithiothreitol, 8% glycerol) supplemented with 0.5% triton, 30 U ml⁻¹ Turbo DNase (Ambion), and 100 µg ml⁻¹ cycloheximide, and ribosome-protected fragments were then generated, cloned and sequenced as previously²³. Briefly, the lysate was cleared by centrifugation, treated with RNase I for 45 min and then loaded on a sucrose cushion. RNA was extracted cushion pellet using TRI reagent and small RNA fragments (28–34 bp) were size selected via gel purification. These RNA fragments were then dephosphorylated, ligated with an adaptor to their 3'-end, and reverse transcribed. The resulting cDNA was circularized and PCR amplified introducing Illumina sequencing adaptors.

RNA-seq and data analysis. Strand-specific mRNA-seq libraries were prepared from U2OS cells using the TruSeq Stranded mRNA Library Prep Kit (Illumina), according to the manufacturer's protocol, and sequenced on a NextSeq 500 machine to obtain at least 23 million 75 nt reads. Strand-specific mRNA-seq libraries for HeLa cells were prepared as described⁴⁰. Briefly, the RNA was fragmented with base hydrolysis and fragments between 26 and 32 nt were gel-extracted. Adaptors containing fixed sequences were ligated to the 3'- and 5'-ends of the RNA fragments followed by additional gel extractions, and after cDNA synthesis, Illumina sequencing adaptors were added using PCR. Reads were aligned to the human genome (hg19 assembly) using STAR Aligner⁴¹, and read counts for individual genes (defined as overlapping sets of RefSeq transcripts annotated with their Entrez Gene identifier) were counted using htseq-count⁴² and normalized to reads per million aligned reads. For counting intron-mapping reads, htseq-count was used to count reads mapping to the whole-gene locus, and the exon-mapping reads were then subtracted for each gene. Only genes with an average RPM of at least 50 normalized reads across the experimental conditions were considered, and fold changes were computed after addition of a pseudo-count of 0.1 to the RPM in

each condition. The raw read counts and the computed fold-changes appear in Supplementary Data 2.

Sequence analyses. Whole-genome alignments were obtained from the UCSC genome browser. Expected numbers of PREs were computed by applying dinucleotide-preserving permutations to the sequences and counting motif occurrences in the shuffled sequences. 3'-UTR-length-matched control targets were selected by dividing the genes into 10 bins based on 3'-UTR lengths and randomly sampling the same numbers of genes not enriched with Pumilio target sites as the number of genes enriched with sites from each bin.

Luciferase assays. Reporter gene activity was measured as previously described⁴³. Briefly, 20,000 cells were plated in a 96-well plate. After 24 h cells were co-transfected with pGL4.13 as an internal control and with the indicated psiCheck plasmids. In addition, the cells were transfected with the various siRNAs or plasmids (as described above). Luciferase activity was recorded 48 h post transfection using the Dual-Glo Luciferase Assay System (Promega) in the Micro plate Luminometer device (Veritas). A relative response ratio, from RnLuc signal/FFLuc signal, was calculated for each sample. Percent of change is relative to the control siRNA or control plasmid.

Determination of copy number of PUM1 and PUM2 in U2OS cells. PUM1 and PUM2 were expressed in bacteria. Briefly, PUM1 and PUM2 cDNA were cloned into a modified version of pMal-C2 expression vector (a kind gift from the laboratory of Prof. Deborah Fass) by restriction free cloning resulting in a MBP-6His-PUM constructs. The plasmids were transformed into Rosetta-R3 bacteria (Novagen). Bacteria were grown in 15 ml 2YT media in the presence of 100 µg ml⁻¹ ampicillin and 50 µg ml⁻¹ chloramphenicol to OD600 ≈ 0.6.

Recombinant protein expression was induced for 18 h at 16 °C by 500 µM IPTG. Bacterial pellet was resuspended in 5 ml of lysis buffer B (100 mM NaH₂PO₄, 10 mM Tris, 8 M Urea, pH8) and incubated on a rotating shaker for 90 min at room temperature. The extract was cleared by centrifugation (10,000g, 20 °C for 30 min). Cleared extract was incubated for 60 min with 1 ml of 50% Nickel beads slurry (Ni-NTA His*Bind Rasin, Novagen) and the extract bead mix was loaded onto an empty column. The column was washed twice with wash buffer C (100 mM NaH₂PO₄, 10 mM Tris, 8 M Urea, pH6.3) and the bound proteins were eluted four times in 500 µl of elution buffer D (100 mM NaH₂PO₄, 10 mM Tris, 8 M Urea, pH5.9), followed by four times in 500 µl of elution buffer E (100 mM NaH₂PO₄, 10 mM Tris, 8 M Urea, pH4.5). Sample of each fraction was run on SDS-polyacrylamide gel electrophoresis and analysed by Coomassie blue staining. To determine the quantity of PUM1 and PUM2 copies per cell we calibrated a standard curve using the purified bacterial expressed PUM proteins and then plotted the protein expression levels from a lysate extracted from a measured number of cells.

Statistics. All results are represented as an average ± s.e.m. of at least three independent experiments. Statistics was performed as Student's *t*-test, Wilcoxon rank-sum test or analysis of variance with Tukey's *post hoc* test for three or more groups to be compared. In all results **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Plots were prepared using custom R scripts. Gene Ontology enrichment analysis was performed using the WebGestalt server⁴⁴ and corrected for multiple testing using TANGO³⁴, using all the expressed genes as background set and Benjamini-Hochberg correction for multiple testing.

Data availability. All data presented in this work is available from the authors upon request. All sequencing data has been deposited to the GEO database (Accession GSE79804).

References

- Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
- Wapinski, O. & Chang, H. Y. Long noncoding RNAs and human disease. *Trends Cell Biol.* **21**, 354–361 (2011).
- Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
- Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).
- Carriero, C. *et al.* Long non-coding antisense RNA controls *Uchl1* translation through an embedded SINEB2 repeat. *Nature* **491**, 454–457 (2012).
- Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284–288 (2011).
- Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
- Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
- Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* **15**, 409 (2014).
- Spassov, D. S. & Jurecic, R. The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life* **55**, 359–366 (2003).
- Weidmann, C. A. *et al.* The RNA binding domain of Pumilio antagonizes poly-adenosine binding protein and accelerates deadenylation. *RNA* **20**, 1298–1319 (2014).
- Parisi, M. & Lin, H. The *Drosophila* *pumilio* gene encodes two functional protein isoforms that play multiple roles in germline development, gonadogenesis, oogenesis and embryogenesis. *Genetics* **153**, 235–250 (1999).
- Menon, K. P. *et al.* The translational repressor Pumilio regulates presynaptic morphology and controls postsynaptic accumulation of translation factor eIF-4E. *Neuron* **44**, 663–676 (2004).
- Galgano, A. *et al.* Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS ONE* **3**, e3164 (2008).
- Vessey, J. P. *et al.* Dendritic localization of the translational repressor Pumilio 2 and its contribution to dendritic stress granules. *J. Neurosci.* **26**, 6496–6508 (2006).
- Lee, M. H. *et al.* Conserved regulation of MAP kinase expression by PUF RNA-binding proteins. *PLoS Genet.* **3**, e233 (2007).
- Moore, F. L. *et al.* Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc. Natl Acad. Sci. USA* **100**, 538–543 (2003).
- Lee, S. *et al.* Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164**, 69–80 (2016).
- Raj, A. *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
- Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110–1122 (2015).
- Wang, L. *et al.* CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74 (2013).
- Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349 (2007).
- Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
- Rubio, C. A. *et al.* Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. *Genome Biol.* **15**, 476 (2014).
- Ingolia, N. T. *et al.* Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* **8**, 1365–1379 (2014).
- Eichhorn, S. W. *et al.* mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol. Cell* **56**, 104–115 (2014).
- Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
- Guttman, M. *et al.* Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013).
- Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.* **2**, e33 (2006).
- Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
- Van Etten, J. *et al.* Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J. Biol. Chem.* **287**, 36370–36383 (2012).
- Gaidatzis, D., Burger, L., Florescu, M. & Stadler, M. B. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat. Biotechnol.* **33**, 722–729 (2015).
- Leeb, M. *et al.* Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* **14**, 385–393 (2014).
- Ulitsky, I. *et al.* Expander: from expression microarrays to networks and functions. *Nat. Protoc.* **5**, 303–322 (2010).
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
- Durocher, Y., Perret, S. & Kamen, A. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res.* **30**, E9 (2002).
- Narita, R. *et al.* A novel function of human Pumilio proteins in cytoplasmic sensing of viral infection. *PLoS Pathog.* **10**, e1004417 (2014).
- Isaac, R. *et al.* Selective serotonin reuptake inhibitors (SSRIs) inhibit insulin secretion and action in pancreatic beta cells. *J. Biol. Chem.* **288**, 5682–5693 (2013).
- Yoon, J. H. *et al.* LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* **47**, 648–655 (2012).

40. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
41. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
42. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
43. Van Etten, J., Schagat, T. L. & Goldstrohm, A. C. A guide to design and optimization of reporter assays for 3' untranslated region mediated regulation of mammalian messenger RNAs. *Methods* **63**, 110–118 (2013).
44. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GENE SeT Analysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
45. The FANTOM5 Consortium. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
46. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res.* **22**, 1173–1183 (2012).
47. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011).
48. Pearson, W. R. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* **132**, 185–219 (2000).
49. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12**, R79 (2011).
50. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

Acknowledgements

We thank members of the Ulitsky laboratory for useful discussions and comments on the manuscript. I.U. is incumbent of the Sygnet Career Development Chair for Bioinformatics and recipient of an Alon Fellowship. Work in the Ulitsky laboratory is supported by grants to I.U. from the European Research Council (Project 'lincSAFARI'), Israeli Science Foundation (1242/14 and 1984/14), the I-CORE Program of the Planning and Budgeting Committee and The Israel Science Foundation (grant no 1796/12), the Minerva Foundation, the Fritz-Thyssen Foundation and by a research grant from The Abramson Family Center for Young Scientists. N.S.-G. is incumbent of the Skirball career development chair in new scientist. S.I. is supported by the Henry Chanoch Krenter Institute for Biomedical Imaging and Genomics, The Leir Charitable Foundations,

Richard Jakubskind Laboratory of Systems Biology, Cymerman-Jakubskind Prize, The Lord Sieff of Brimpton Memorial Fund, the Human Frontiers Science Program, the I-CORE program of the Planning and Budgeting Committee and the Israel Science Foundation, the European Molecular Biology Organization Young Investigator Program and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement number 335122. S.I. is the incumbent of the Philip Harris and Gerald Ronson Career Development Chair.

Author contributions

A.T. and I.U. conceived the study. A.T. designed and performed most experiments with input from the other authors. N.G. analysed the PAR-CLIP data. Y.L. performed the protein copy number quantification experiments. T.H.S. and N.S.G. performed the ribosome footprinting experiments. D.L. and S.I. analysed the smFISH data. I.U. analysed the RNA-seq and Ribo-seq data. All authors contributed to writing of the paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Tichon, A. *et al.* A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**:12209 doi: 10.1038/ncomms12209 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016